

## Language Technology II Natural-language Interaction

Manfred Pinkal  
Magdalena Wolska

Course website:  
[www.coli.uni-saarland.de/courses/late2](http://www.coli.uni-saarland.de/courses/late2)



## Dialogue System

A dialogue system is a computer system whose purpose is to **achieve an application goal** through a series of **interactions in natural language** with a user.

Language Technology II, Summer 2008 © M. Pinkal



## Talking Elevator

- User: Elevator
- System: Which floor do you want to go?
- User: Four /Fourth floor
- System: <Command execution>

Language Technology II, Summer 2008 © M. Pinkal



## Plausible Scenarios for NL Interaction

- Device operation for disabled people
- “Hands-on” (and “eyes-on”) situations
  - Car-driving, Maintenance, “Pick-by-voice” in logistics
- Telephony applications:
  - travel arrangements, customer service, telephone directory, ...

Language Technology II, Summer 2008 © M. Pinkal



## Dialogue System: Possible Goals

A dialogue system is a computer system whose purpose is to **achieve an application goal** through a series of **interactions in natural language** with a user

Possible Goals:

- Retrieve information: **Information-seeking dialogue**
- Seek to satisfy constraints: **Negotiation dialogue**
- Perform action: **Command&control dialogue**
- Collaborate on solving a problem: **Problem-solving dialogue**
- Instruct: Help/ Instructional/ **Tutorial dialogue**



## A Tutorial Dialogue

S: I didn't understand. What force acts on all objects within close proximity to earth?  
 U: *uh the force of gravity*  
 S: Fine. Besides the Earth's gravitational force, are there any other forces acting on the person, keys, and elevator after the release?  
 U: *no*  
 S: Yep. Thus, keep in mind that after the release, the only force on the person, keys, and elevator is the force of gravity. If gravity is the only force acting on an object, will it be moving or staying still?  
 U: *it will be moving*  
 S: Fine. Not only are the person, keys, and elevator moving, they have only gravitational forces on them. When an object is falling and has only gravitational force on it, it is said to be in what?  
 U: *freefall*

**Figure 19.3** The tutoring domain: part of an actual conversation between a student user (U) and the ITSPOKE system (S) of (Litman and Silliman, 2004), based on the Why2-Atlas text-based tutoring system (?).



## Travel Arrangement Dialogue

S1: Hello. You've reached the [Communicator]. Tell me your full name  
 U2: *hi i'd like to fly to seattle tuesday morning*  
 S3: Traveling to Seattle on Tuesday, August eleventh in the morning. Your full name?  
 U4: *[John Doe]*  
 S5: Hello, [John], I'm glad to hear from you again. You can now ask me for help.  
 S6: OK. Let's see I have a United flight departing Pittsburgh at ten forty five a.m. arriving into Seattle at one fifty seven p.m. Is that OK?  
 U7: *Yes*  
 S8: Will you return to Pittsburgh from Seattle?  
 U9: *no i will go to Los Angeles*  
 S10: Traveling to Los Angeles. on what date  
 U11: *Sorry can we go back to talk about the previous leg again*  
 S12: For the first leg, you are on a United flight 1446 departing Pittsburgh at ten forty five A.M. tomorrow, arriving in Seattle at one fifty seven P.M. Do you want to take that?  
 U13: *actually i want to leave on wednesday*  
 S14: a flight on wednesday.  
 S15: I have a flight departing Pittsburgh at ten forty five a.m., arrives Seattle at one fifty seven p.m. Is that OK?  
 U16: *Yes*

**Figure 19.1** The travel domain: a fragment from a successful conversation between a user (U) and the Communicator system (S) of Xu and Rudnicky (2000).



## Further Dialogue Classification

- Interaction Style:
  - System-directed
  - User-directed
  - Mixed-initiative



## Further Dialogue Classification

- Interaction Style:
  - System-directed
  - User-directed
  - Mixed-initiative
- Modality:
  - Monomodal, spoken language
  - Multimodal / crossmodal dialogue, combining two or more modalities

Language Technology II, Summer 2008 © M. Pinkal



## Further Dialogue Classification

- Interaction Style:
  - System-directed
  - User-directed
  - Mixed-initiative
- Modality:
  - Monomodal, spoken language
  - Multimodal / crossmodal dialogue, combining two or more modalities:
    - Keyboard, buttons, pointing devices, graphics
    - More advanced: Gesture, Facial Expression

Language Technology II, Summer 2008 © M. Pinkal



## Dialogue Systems: State of Technology

- Various academic systems with impressive demo level performance are available
- Commercially deployed systems:
  - Telephony
  - Simple Command&Control device operation
  - Dictation systems
- User acceptance of all kinds of systems is low
- Why this gap between science and application?
  - Unsatisfactory speech recognition
  - Poor dialogue design
  - Users' expectations are tremendously high: experience with other technical systems or their experience with human dialog partners

Language Technology II, Summer 2008 © M. Pinkal



## Dialogue Design

© 1999 Randy Glasbergen.  
www.glasbergen.com



**“...If you'd like to hear all of your options again, press 49. If you've forgotten why you called in the first place, press 50.”**

Language Technology II, Summer 2008 © M. Pinkal



## Talking Elevator

- User: Elevator
- System: Which floor do you want to go?
- User: Fourth floor
- System: So you want to go to the fourth floor?
- User: Yes, please.
- System: <Command execution>

Language Technology II, Summer 2008 © M. Pinkal



## Talking Elevator

- User: Elevator
- System: Which floor do you want to go?
- User: Fourth floor
- System: I'll take you to the fourth floor.  
<Command execution>

Language Technology II, Summer 2008 © M. Pinkal



## Talking Elevator

- User: Elevator
- System: Which floor do you want to go?
- User: Fourth floor
- System: I'll take you to the fifth floor.
- User: No, fourth floor!
- System: Ok, I'll take you to the fourth floor.  
<Command execution>

(Explicit or implicit) Grounding: Triggers correction by user.

Language Technology II, Summer 2008 © M. Pinkal



## First Generation Dialogue Systems

- "Command and Control" paradigm
- Restricted vocabulary, keywords, system-initiated and controlled rigid dialogue structure
- Optimising for robustness and effectiveness (to make good for poor speech recognition quality)

Language Technology II, Summer 2008 © M. Pinkal



## Second Generation Dialogue Systems

- Flexible, adaptive, cooperative dialogue systems
- Optimising for efficiency and user satisfaction
- **Flexible and content orientated:** Users say *what* they want, and *the way* they want, with no need to learn the specific commands a device “expects”: Lexical choice, reference identification, syntax
- **Collaborative:** Clarification of ambiguous or underspecified user input.
- **Adaptive:** Adjusting to user types (novice, expert), also dynamic adaptation to situation (driving!) and user’s learning curve.



## Talking Elevator

- User: Elevator
- System: Which floor do you want to go?
- User: To professor Barry's office.
- System: I'll take you to the fourth floor.  
<Command execution>



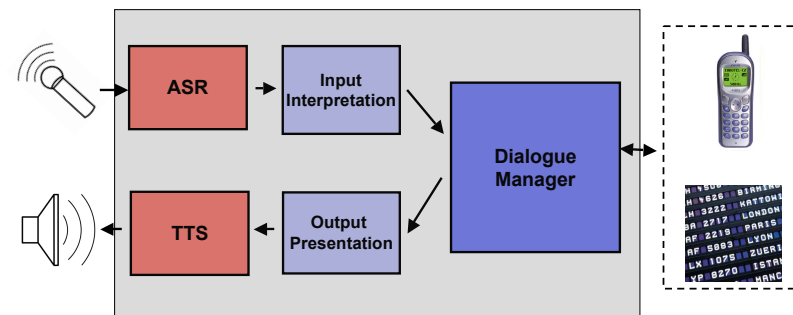
## Talking Elevator

- User: Elevator
- System: Which floor do you want to go?
- User: To the phonetics department.
- System: Fourth or fifth floor?
- User: Professor Barry's office.
- System: I'll take you to the fourth floor.  
<Command execution>

Clarification dialogue, system presenting a contextual determined alternative.



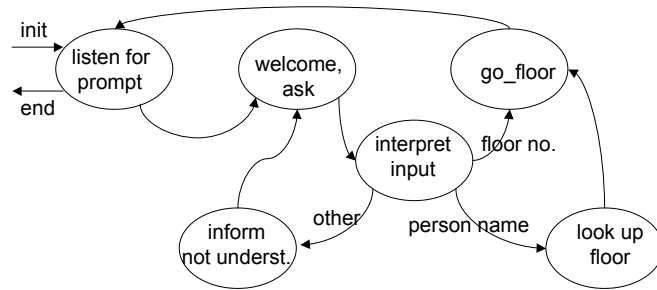
## Basic Architecture of a Dialogue System





## Dialogue Management

- Finite state diagram models



Language Technology II, Summer 2008 © M. Pinkal



## Dialogue Management

- Frame-based or form-filling dialog models

[SHOW:  
 FLIGHTS:  
   (getOrigin CITY)  
   (getDate DATE) (getTime TIME)  
 DEST:  
   (getDestination CITY)]

U1: Show me flights to SF.  
 U2: Show me morning flights from Boston to SF on Tuesday.

Language Technology II, Summer 2008 © M. Pinkal



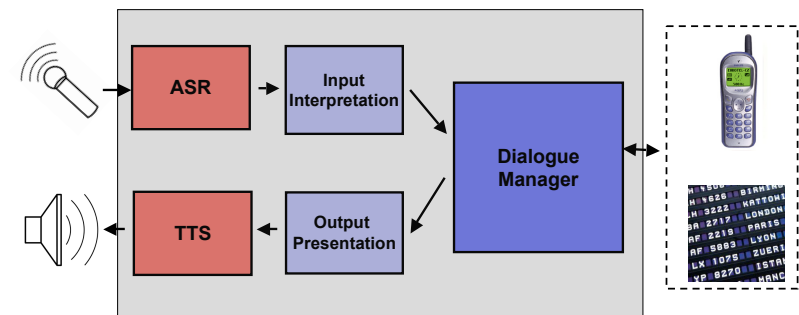
## Dialogue Management

- Finite State Diagrams
- Frame-Based /Form-Filling Dialogue Models
- Information State Update
- Plan-Based Dialogue Systems

Language Technology II, Summer 2008 © M. Pinkal



## Basic Architecture of a Dialogue System



Language Technology II, Summer 2008 © M. Pinkal



## Automatic Speech Recognition

- Speech Signal Recording and Pre-Processing
- Spectral Analysis and Feature Extraction

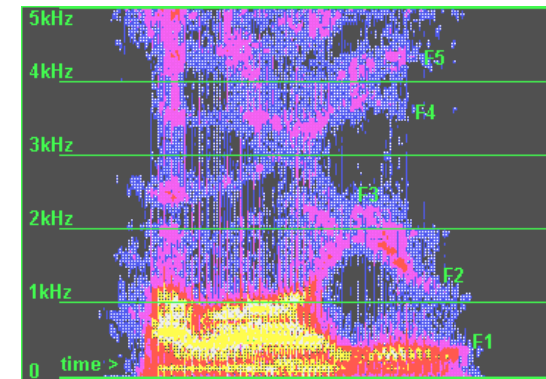
Speech Recognizer, using HMM Technology

- Acoustic Models: Relate Speech Signal Information to Phonemes and Words
- Language Models: Provide information of which word is might come next in the context of a string of words already recognized

Language Technology II, Summer 2008 © M. Pinkal



## A Spectrogram



How are you?

Language Technology II, Summer 2008 © M. Pinkal



## Automatic Speech Recognition

Why acoustic models **plus** language models?

- Recognition quality heavily depends on the range of possible alternatives (more general: on “perplexity”)
- Acoustic information is in principle insufficient for word recognition
  - *Heutzutage fahren immer mehr ans Meer*
  - *Heutzutage fahren immer Meer ans mehr*
  - *From this point, you can see the sea*
  - *From this point, you can sea the see*

Language Technology II, Summer 2008 © M. Pinkal



## Automatic Speech Recognition

How do we get at language models?

Two basic methods:

- Hand-coded Recognition Grammars
- Statistical Language Models (SLMs)

Language Technology II, Summer 2008 © M. Pinkal



## Recognition Grammars

- Hand-coded Recognition Grammars
  - Typically written in EBNF notation ( Context-free grammars)
  - Typically shallow “semantic grammars” with no recursion
  - Are compiled to regular grammars/finite automata (by ASR system) without loss of information
- An example:
 

```
$turn = [please] turn | turn $direction ;
$direction= (back|backward)| $side;
$side = [to the](left | right)
```

Language Technology II, Summer 2008 © M. Pinkal



## Properties of recognition grammars

- Allow quick and easy specification of application-specific and dialogue-state specific language models
- Thereby enormously reduce search space/ perplexity for recognizer
  - Example: \$yn\_answer = yes | no
- But: Strictly constrain recognition results to the language specified in the grammar.

Language Technology II, Summer 2008 © M. Pinkal



## Properties of recognition grammars

- Option 1: Keyword Spotting
  - Working with wildcards
 Example:
 

```
$turn = GARBAGE* turn | turn $direction GARBAGE* ;
$direction= (back|backward)| $side;
$side = GARBAGE* (left | right)
```

  - No relevant lexical information is lost, but recogniser performance decreases
- Option 2: Including all variants and semantically irrelevant material into the grammar
  - Makes recognition grammar writing cumbersome, and will never be complete.

Language Technology II, Summer 2008 © M. Pinkal



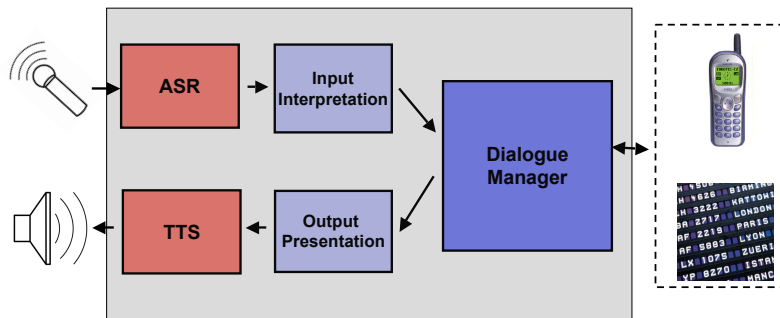
## Statistical Language Models

- Statistical Language Models (SLMs) are
  - Trained on text or transliterated dialogue corpora
  - Based on n-gram (typically trigram) probabilities
- If trained on really large text corpora, SLMs provide high recognition quality (dictation systems)
- Training of SLMs for dialogue needs domain- and application-specific dialogue corpora, which is expensive:
  - Data collection in Wizard-of-Oz experiments
  - Manual transliteration

Language Technology II, Summer 2008 © M. Pinkal



## Basic Architecture of a Dialogue System



Language Technology II, Summer 2008 © M. Pinkal



## Input Interpretation

- It is useful to specify dialogue models not with respect to utterances or key words, but to semantic concepts
  - „5:15 p.m.“, „17:15“ „a quarter past five“ express the same content
- How is input interpreted?

Language Technology II, Summer 2008 © M. Pinkal



## Recognition Grammars

- Recognition grammars with **interpretation tags** have double function. They constrain the language model and interpret the recognised input.
- An example:
 

```
$turn = [please] turn {$.action="turn"}
  | turn $direction {$.direction=$direction} {$.action="turn"};
$.direction= (back|backward) {"backward"} | $side {$=$side};
$.side = [to the](left {"left"} | right {"right"})
```

Language Technology II, Summer 2008 © M. Pinkal



## Interpretation for SLMs

Interpretation Grammars:

- Interpretation grammars look very similar to recognition grammars.
- But they work completely different : They parse the speech recogniser output (typically on the best chain)
- SLMs are permissive with respect to the material they (in part erroneously) hypothesise.
- Therefore permissive parsers are needed, which may skip material (and assign a penalty).
- An example: An early parser building up a chart, and selecting the best path (w.r.to the number of omitted words).

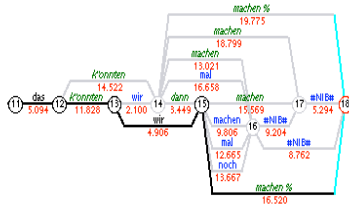
A completely different, but promising option: Concept recognition

Language Technology II, Summer 2008 © M. Pinkal



## ASR Output

- Word Lattice (Word Hypothesis Graph)



- n-Best: List of word chains read off from paths through the word lattice, ordered by probability
- Confidence values on turn or word level.

Language Technology II, Summer 2008 © M. Pinkal



## Performance of ASR Systems

- Standard performance measure is "Word Error Rate" (WER):
  - The **Minimum Edit Distance** between best hypothesis and correct string: Insertions+Substitutions+Deletions
  - divided by total number of words in correct string.
  - Example:
    - Yes, *that would be a good idea. Lets just do it.*
    - Yes, *then* would be a good idea. Lets just do.
- „Concept error rate“ might provide a more reliable measure.

Language Technology II, Summer 2008 © M. Pinkal



## Performance of ASR Systems

How good are state-of-the-art ASR systems?

Performance is dependent on :

- Acoustic quality of input signal
  - recording and transmission (microphone, telephone, mobile phone)
  - background noise (especially distraction by speech)
  - room acoustics
- Preprocessing (e.g., noise cancellation)
- Phonetic and linguistic quality of input utterance (voice, accent, spontaneous speech)
- The task: Lexicon size and perplexity
- The recogniser itself: Dependent on the available platform (server/ standard PC / embedded)

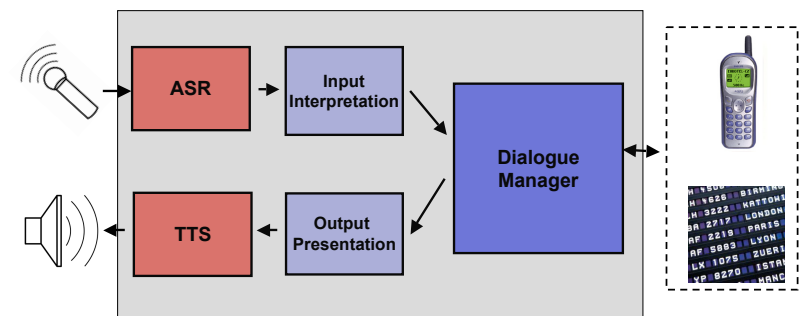
Challenges:

- Very large vocabulary (city and street names for navigation)
- Multi-slot input (city+street name in one turn)

Language Technology II, Summer 2008 © M. Pinkal



## Basic Architecture of a Dialogue System



Language Technology II, Summer 2008 © M. Pinkal



## Synthesis: Text-to-Speech

- Diphone synthesis
  - Poor quality, bad acceptance
- Pre-recorded speech
  - High quality, but expensive and risky
- Current Standard Technology: Unit selection
  - Speech output is combined from a database containing soundfiles of different size (Phonemes, Syllables, Words, Phrases)
- A challenge: Prosody
  - Accent requires semantic contextual information (information structure): „concept to speech“
  - Commercially deployed systems provide „neutral“ pronunciation, including moderate question intonation.

Language Technology II, Summer 2008 © M. Pinkal



## Generation / Output Presentation

- Template-based generation:
  - The next flight to **\$AIRPORT** will leave at **\$DAYTIME**.

Language Technology II, Summer 2008 © M. Pinkal



## NL Interaction: Classes

- 20.06 : Introduction and Overview
- 27.06 : Dialogue Phenomena
- 03.06 : Dialogue Modelling
- 10.06 : Empirical techniques in dialogue systems
- 17.06 : Multilingual aspects of dialogue

Language Technology II, Summer 2008 © M. Pinkal